Brochure

HP ProLiant Gen8と Red Hatで始める Hadoop™

Hadoop™スタートアップ支援サービス

「ビッグデータ」から成長への価値を引き出すには、Hadoopをいかに使いこなすか、が決め手です。

企業が競争に勝利するうえで、ますます重要度を増している「ビッグデータ」。
この新しいタイプのデータから成功への価値を引き出すには、
まったく新しいテクノロジーが必要であり、Hadoopはその最有力の選択肢です。
Hadoop活用の基盤となるシステムをいかに実現するか。最適な回答へたどり着けるよう、
HPはHadoopシステム構築に向けた実践的な支援サービスを提供しています。

急がれるビッグデータへの対応

ビッグデータは、IT業界における比較的新しい用語です。増大するビッグデータは、多くの企業内に蓄積されている膨大な生データというにとどまらず、リッチデータが標準となっている業界において、すでに象徴的な課題となっています。

ビッグデータを無視することはもはや許されません。IDC社では、デジタルコンテンツの容量が2012年には2.7ZBに増大し（2011年の48%増）、2015年までに8ZBに達すると予測しており、モバイルおよびクラウドに加えてビッグデータが、競争力を維持するうえで「必須の」要素になると見ています2。適切な管理手法を確立できなければ、企業はこの膨大なデータに圧倒されてしまいます。

しかし、ビッグデータは困難をもたらすだけのものではありません。新しいタイプのデータを活用することで、企業はアジリティ（俊敏性）を向上させることができ、従来は不可能であった要求に応えることが可能になります。ビッグデータ手法はこれまでにない革新的な手法であり、ビジネスにおける既存の問題の解決、新規ビジネスモデルの実装、および成長の促進にも貢献します。

IDC社では、デジタルコンテンツの容量が2012年には2.7ZBに増大し（2011年の48%増）、2015年までに8ZBに達すると予測しており、モバイルおよびクラウドに加えてビッグデータが、競争力を維持するうえで「必須の」要素になると見ています。

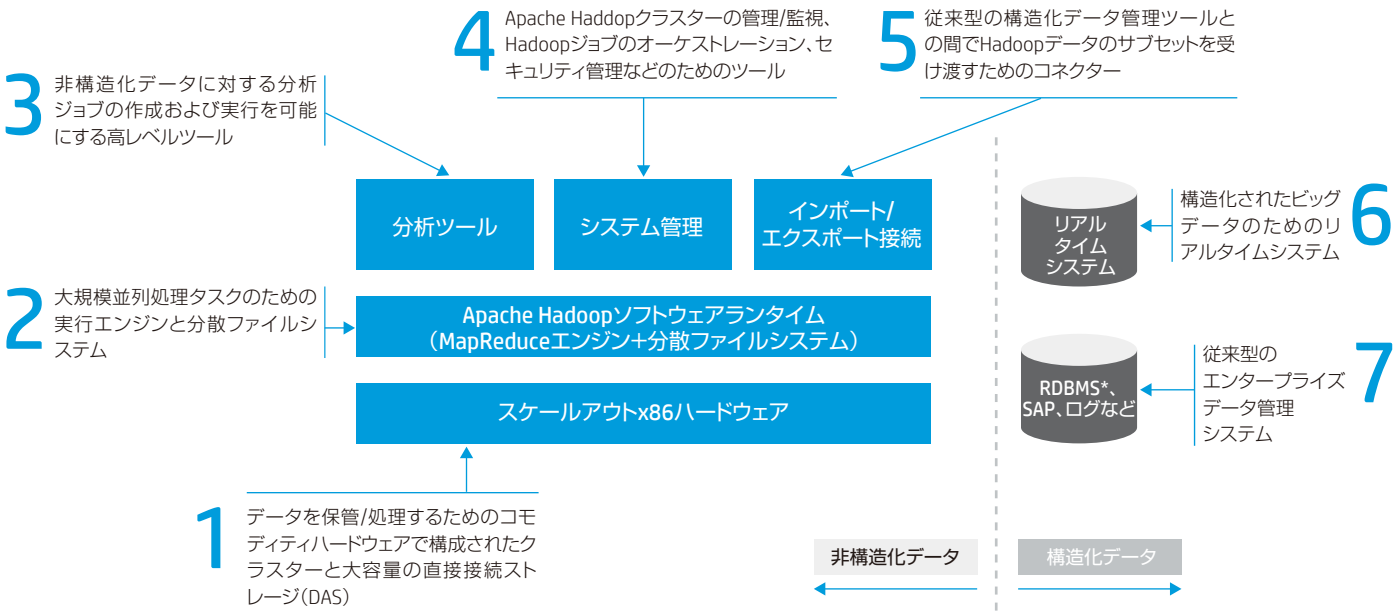
※1「DC Predictions 2012: Competing for 2020」、IDC社、2011年12月。
注記: 1ZB =10億テラバイト

ビッグデータ活用に道を拓く最良のテクノロジー Hadoop

ビッグデータから価値を引き出すために、企業はビッグデータを保管、処理、および分析するための新たなテクノロジーを必要としています。保有

■Hadoopエコシステムの構成要素

Hadoopエコシステムは、非構造化データに対するバッチベースの大規模並列ジョブを実行するための体系であり、多数の先進的なビッグデータソフトウェアツール、および従来の構造化データ管理ツールとのインターフェイスで構成されています。



* Relational Database Management System（リレーショナルデータベース管理システム）の略

するビジネスデータの価値は、ビッグデータのためのオープンソースフレームワークとして新たに登場したApache Hadoopは、費用効果が高く革新的なアプローチによる高度な分析を通じて、新たな可能性を企業に提供します。Hadoopを使用することで、非構造化データを低コストで保管し、妥当な時間内に処理できるため、データ集約型のアプリケーションを費用効果の高いハードウェア上で実行することが可能になります。
多くの企業が、Hadoopの拡張性の高さと、低コストのコンポーネントを使用して大容量データを処理できる点に注目しています。データの保管と処理に伴うコストを抑制可能なHadoopは、非構造化データの活用を目指している企業に魅力的な選択肢を提供します。さらにHadoopでは、既存のビジネスインテリジェンスシステムでの使用に適した形で、圧縮された構造化サマリーを提供することも可能です。

拡張性、パフォーマンス、および費用効率を基本原則として構築されたHadoopは、Javaベースのオープンソースフレームワークとして、低コストのスケールアウトアーキテクチャー上での大規模並列データ処理プロジェクトの運用を可能にします。

優れたコスト効率と高い“弾力性”、耐障害性を特徴とするHadoopシステム

Hadoopはクラスター化されたマシンを活用し、それらのマシン間で処理を調整します。クラスターの構築および拡張には、汎用のコンピューターを使用できます。Hadoopソフトウェアパッケージの主な構成要素は、以下のとおりです。

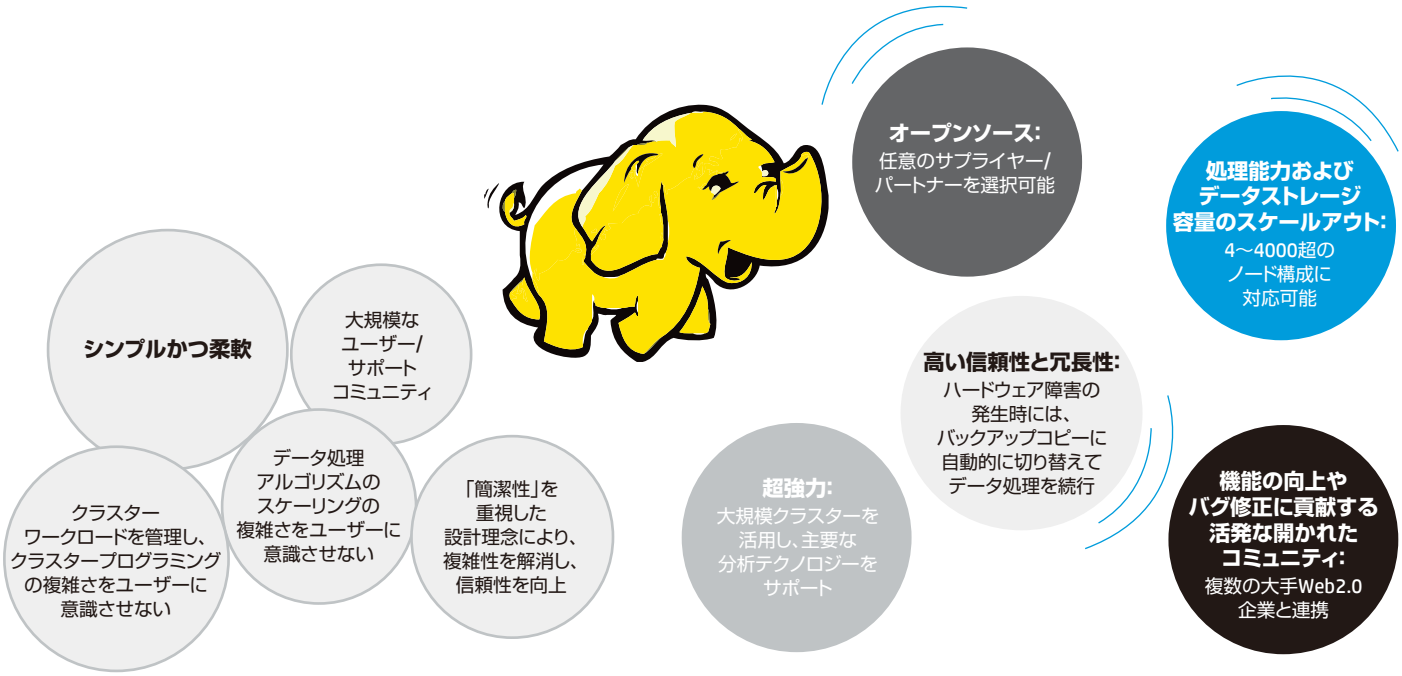
- **Hadoop分散ファイルシステム（HDFS: Hadoop Distributed File System）**: 個々のデータ処理ノード内に存在する多数の汎用ディスクドライブ（SATAドライブなど）で構成される分散型ストレージ容量を使用して、低コストの冗長ファイルシステムを提供する大規模データ並列ファイルシステムです。
- **MapReduce**: 非構造化データ内の情報を抽出、変換、エクスポートするための並列プログラミングフレームワークです。

Hadoopインフラストラクチャは「弾力性」を特徴とし、新たなデータノードを追加することでネットワークやディスクの能力および帯域幅を簡単に拡張できます。またHadoopフレームワークによるスケジュールと管理により、データソースに最も近いノードでデータが処理されることで、最大限のパフォーマンスが達成されます。HDFSは物理クラスターアーキテクチャーも認識可能で、クラスターにわたってデータを複製することで耐障害性の強化にも貢献します。また、Hadoopにより管理されているいずれかの複製から追加のデータコピーが自動的に作成されます。このように、Hadoopクラスターはストレージおよび処理能力のいずれについても自己修復機能を備えているため、システム管理者による頻繁な介入は必要ありません。

最新のHP ProLiant Gen8をベースにHPの知見を投入して解決策をご提案

最新の技術であるHadoopは、世界的に見てもその適用が立ち上がったばかりであり、目的に応じた最適なシステム構成、分散ファイルシステムの実装法、パラメーター設定、運用ポイントといった実務に則した情報は極めて限られています。このため、従来型のデータウェアハウス手法をHadoopに適用するという過ちを招いてしまうこともあります。
HPでは、特に困難が伴うスタートアップ時の課題解消を支援すべく、社内での技術検証やワールドワイドでのノウハウ蓄積をいち早く実施。“自働サーバー”として注目を集める最新の汎用サーバーHP ProLiant Gen8を核として、豊富なベストプラクティスを積み上げています。Hadoopを最短期間で導入し、最大限の効果を引き出すために、HPの広範なHadoopスタートアップ支援サービスをぜひ活用ください。

■Hadoopの主な特徴



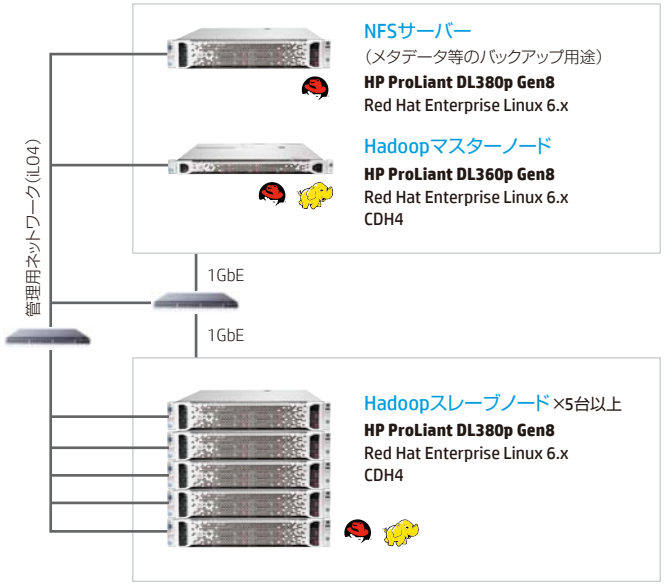
HPが自らの検証成果を基に自信をもって組み上げた
オススメの構成例だから確実に、スピーディな導入が可能

Hadoopへの注目度は高まっているものの、国内ではまだまだ導入ケースの少ないのが現状です。
そこでHPでは、自ら実施した様々な検証成果を基に、確実で失敗のないオススメ構成例を用意しました。
ハードウェアには汎用サーバーとして定評のあるHP ProLiant Gen8、
OSには大規模なスケールアウトにも対応するRed Hat Enterprise Linux、
そして商用Hadoopとして米国で多くの採用実績を持つCloudera's Distribution including Apache Hadoop（CDH4）。
この3つをベースにしたシステム構成なら、効率的なHadoop導入が可能です。

オススメ構成例1　開発用スモールスタート

システム構成上のポイント

- Hadoopの本格活用に向けた検討をすぐに始めたい、Hadoop活用によるビジネス上のメリットを見極めたい、といったお客様に最適
- 必要最小の構成にすることで初期コストを抑制
- 本格導入後の拡張にも余裕で応えるパフォーマンスとスケーラビリティ
- iLO4による効率的な管理が可能



HP ProLiant Gen8サーバー・ハードウェア基本構成

NFSサーバー

HP ProLiant DL380p Gen8	
CPU	6コア Intel E5-2630 2.3GHz ×2
RAID	P420i Smart Array
HDD	1TB SFF SAS 7.2krpm ×16
メモリー	64GB DDR3
NIC	1GbEポート ×4

Hadoopマスターノード

HP ProLiant DL360p Gen8	
CPU	6コア Intel E5-2630 2.3GHz ×2
RAID	P420i Smart Array
HDD	1TB SFF SAS 7.2krpm ×4
メモリー	64GB DDR3
NIC	1GbEポート ×4

Hadoopスレーブノード

HP ProLiant DL380p Gen8	
CPU	6コア Intel E5-2630 2.3GHz ×2
RAID	P420i Smart Array
HDD	1TB SFF SAS 7.2krpm ×16
メモリー	64GB DDR3
NIC	1GbEポート ×4

Hadoopの効率運用を支える高性能"自働サーバー"、HP ProLiant Gen8

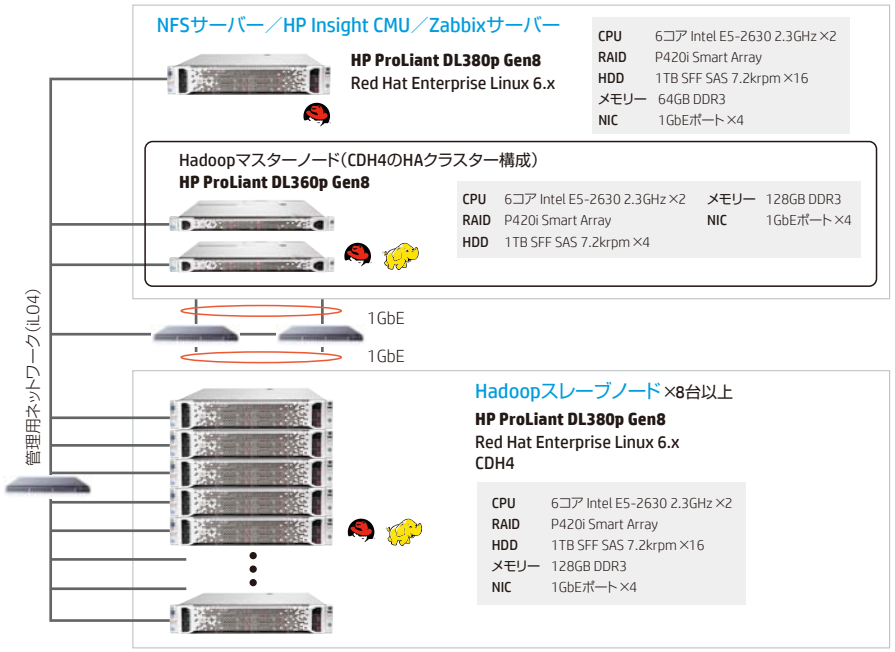
これまで世界中で数多くの採用実績を積み上げてきたコストパフォーマンスに優れるHP ProLiantサーバー。その最新世代機であるHP ProLiant Gen8は、高性能なインテル® Xeon® プロセッサー E5ファミリーを搭載し、ライフサイクル全体を通して工数削減に寄与する大胆な自動化を導入することで、"自働サーバー"へと進化しました。複数のサーバーが連携するHadoopシステムにおいても、HP ProLiant Gen8が備える異次元の管理性は、効率的な運用と安定した稼働に大きく貢献します。

●オススメ構成例で採用している主要なテクノロジーの詳細は表4を参照ください。

オススメ構成例2　本番用スモールスタート

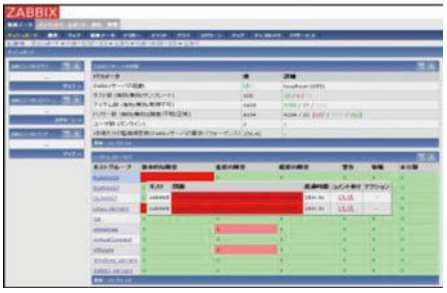
システム構成上のポイント

- ビジネスのためにビッグデータ活用をすぐにでも始めたいというお客様に最適
- Hadoop利用に必要十分な最小構成で初期コストを抑制
- 将来のビジネス成長を見据え、十分なパフォーマンスとスケーラビリティを確保
- iLO4による効率的な管理が可能
- 強力な運用管理ツールが日常的な運用をサポート、運用負担を軽減



運用管理環境

HP Insight Cluster Management Utility（CMU）

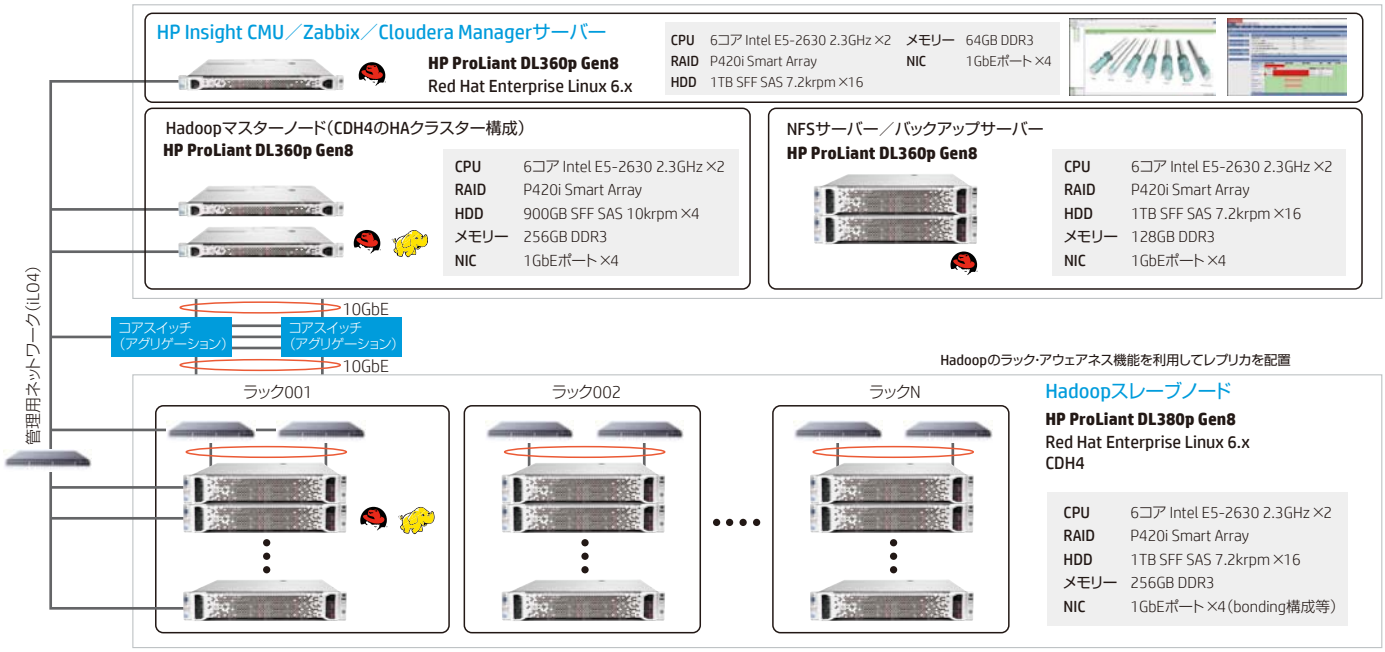


Zabbix for HP ProLiant Gen8 Agentless Management

オススメ構成例3　本番用フルスペック

システム構成上のポイント

- 今あるビッグデータを最大限に活用したい、システムをバッチ処理短縮にも使いたい、といったお客様に最適
- 膨大なデータを収容する本格的な分散ファイルシステムの構築が可能
- ネットワークに10Gbイーサネットを採用し、高いレスポンスを確保
- データの保全性を高めるバックアップ環境も整備
- iLO4、および運用管理ツールにより効率的で作業負荷の少ない運用管理環境を実現



情報が不足しがちなHadoopだから、頼れる案内役が不可欠、HPは計画から構築、運用まで充実した支援サービスを提供します

Hadoopの導入に向けて具体的な取り組みを進めていこうとすると、様々な障害が待受けています。
ビッグデータをどのようにビジネスに活かすのか、目的に最適なHadoopシステムはどのようなものか、
Hadoopの実装に伴うリスクや課題とは、運用上の注意点とは……。
情報が不足している際先端の分野だけに、誤解や判断ミスは投資額の膨張や導入の大幅な遅延といった事態を招きかねません。
そうした困難を回避できるよう、HPでは広範な技術検証などをベースに、
Hadoop活用に関連した多彩な支援サービスを実務的なノウハウを含めて提供。
頼れる案内役として、お客様のプロジェクト成功に必ず貢献できます。

Hadoop導入コンサルティングサービス
導入に関わるテクニカルな課題を解消

サービス概要
Hadoop導入に伴うシステム検討から設計、そして運用まで、技術的なテーマを中心に全工程にわたってコンサルティング。実務的なノウハウに踏み込んで、アドバイスを提供します。

お客様のメリット
- 導入の全工程にわたってワンストップで課題の解消が可能
- 具体的な技術ノウハウを入手でき、スムーズなHadoop導入ができる
- パートナー様／SIer様のソリューションとのHadoop連携にも対応

■Hadoop活用の各フェーズで生まれる課題

検討	設計	開発	テスト	運用
適用領域の選定が困難	Hadoopに関する知識、経験が不足			確立された運用監視方式の欠如
導入効果の可視化が困難	Hadoopkだけでなく、関連/連携するソフトウェアのノウハウが不足			オープンソースに対する保守対応がない

Hadoop HBase関連サービス
効果的なバリューストア構築をサポート

サービス概要
Hadoop上に構築する高性能バリューストアであるHBaseに関して、システム検討から運用まで、全工程にわたってコンサルティングを提供します。

お客様のメリット
- まだ国内では構築事例の少ないHadoop HBaseの導入ノウハウが入手できる
- サーバー、データベース、運用ツールまでトータルな支援が受けられる
- 事前検証やチューニングなども技術サポートも利用できる

■サービスメニュー

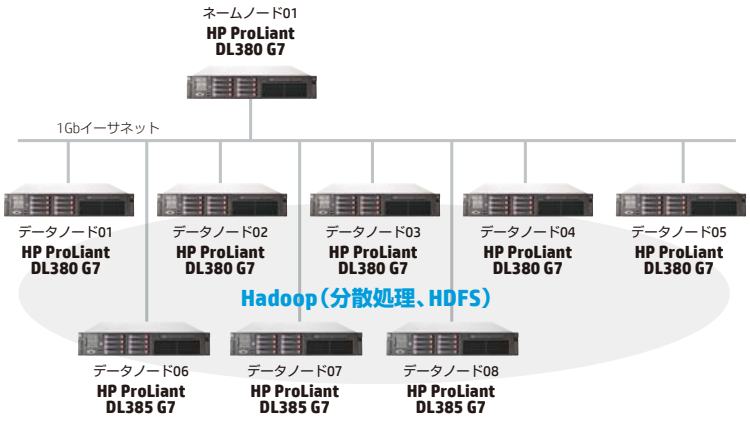
Hadoop HBase コンサルティングサービス
適用領域、既存システムの連携についてのヒアリングを実施
必要に応じて、導入前検証をHPソリューションセンターで実施

Hadoop HBase導入支援サービス
ヒアリング内容に基づくサーバ構成の検討・設計
Hadoopに対応した全体アーキテクチャの設計・構築
性能測定・チューニング

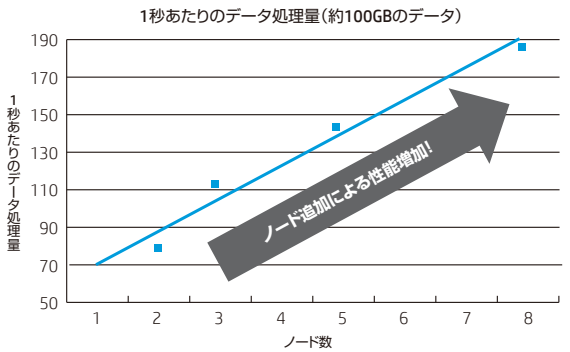
Hadoop HBase運用支援サービス
HP CMUおよびZabbix/Gangliaを使ったクラスター運用基盤の構築
Hadoopクラスター運用支援

HPの技術検証成果1　バッチ処理性能向上検証

■検証環境構成概要図



■検証結果



ノードを追加していくことで、バッチ処理のスピードが向上していくことを確認できた。

HP ビッグデータ分析コンサルティングサービス
経営と一体になったビッグデータ活用を支援

サービス概要
データの抽出法、データ分析の具体的手法などビッグデータを経営に活かし、Hadoopで業務改善や新サービス・商品の開発につなげるためのポイントをお客様と一緒に考えます。

お客様のメリット
- ビッグデータから経営に貢献できる未知の価値を引き出すことが可能に
- 経営サイドに対しビッグデータ活用のアイディア提案が可能に
- ビジネステーマに則したHadoopシステムの構築を実現

■サービスメニュー

1. ディスカバリーワークショップ
お客様の業務、システム改善を目的に、最新製品およびテクノロジーを紹介
統計分析によるテキスト/ログなどの事例紹介およびデモの実施

2. 分析支援コンサルティングサービス
お客様の業務やシステムの全体像の把握、PDCAを回す全体設計
今後の対応策、予算規模、スケジュールなど上申書の基となる情報を作成
システムの将来像とロードマップの策定および全体の参考価格も算出

3. 分析システム導入支援サービス
策定したシステムの将来像に向けて、システム全体の要件定義、システム設計、システム構築サービスを提供
事前検証を実施

4. 分析システム運用支援サービス
システム運用時の支援を実施

HP ビッグデータバッチ処理高速化コンサルティングサービス
低コストで高速なバッチ処理にHadoopを活用

サービス概要
バッチ処理の高速化にも威力を発揮するHadoop。低コストなコモディティサーバーを利用して、容易な変更や拡張が可能なバッチ処理システムの構築をサポートします。

お客様のメリット
- バッチ処理のためのシステムコストを削減
- 柔軟な拡張性を備えたバッチ処理システムを構築可能

■サービスメニュー

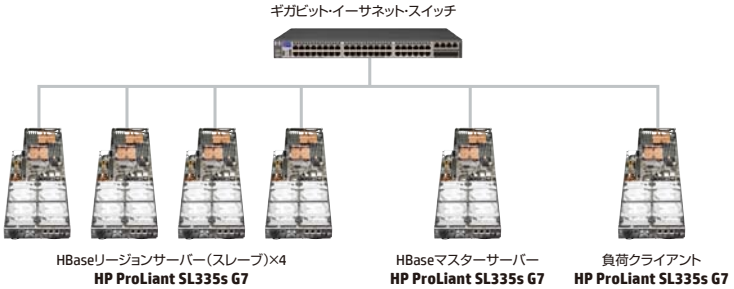
1. バッチ処理高速化コンサルティングサービス
お客様にもアセスメントに参加いただき、既存システムの問題点を的確に把握

2. バッチ処理高速化システム導入支援サービス
アーキテクチャー、運用、監視などのシステム設計を実施
機能テストや、性能テスト、負荷テスト、耐久テストなどを実施
顧客毎のシステム要件に従いプログラミングを日本HPで行う

3. バッチ処理高速化システム運用支援サービス
システム運用時の支援を実施

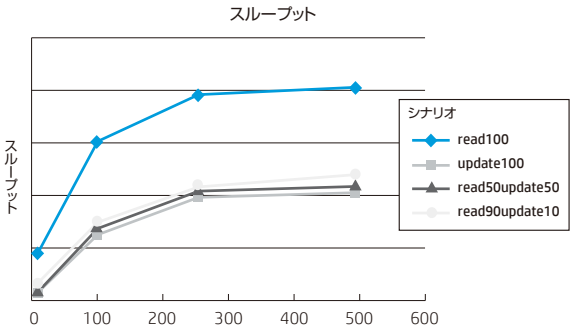
HPの技術検証成果2　HBaseスループット検証

■検証環境構成概要図



試験項目		
	データサイズ	1レコード1KB×3,000,000レコード=3GB
	シナリオ	シナリオ1～シナリオ4
	スレッド数	10、100、250、500、1000の5種類 ※各操作と次の操作の間の待ち時間は0秒

■検証結果



処理のスレッド数を増やしても、HBaseのスループットは頭打ちになってしまうことが判明した。

オススメ構成例で採用したキーテクノロジー

キーテクノロジー1

HP ProLiant DL380p Gen8

幅広い用途に適応する余裕の拡張性とパフォーマンス、
運用効率も劇的に改善する汎用2Uサーバー

- 最大16コア/32スレッドのインテル® Xeon® プロセッサーE5-2600製品ファミリーを搭載。標準24DIMMスロット、4x1GbEまたは2x10GbEのNICを装備し、内蔵USBポートと内部SDスロット装備で仮想化に最適
- 最大25台の2.5型ハードディスクを搭載可能、外付けストレージ要らずの内部拡張性。PCIも最大6スロットまで拡張可能
- iLOマネジメントエンジン(iLO4)がサーバーのライフサイクル管理を自動化
- 「センサーの海」により、冷やすべき箇所だけ冷やすスマートな冷却を実現、独自の省電力機能を継承しつつさらに高性能となった電源を装備し、無駄な電力消費を徹底して排除

キーテクノロジー2

HP ProLiant DL360p Gen8

管理性、省電力性、拡張性にも優れ、
設置スペースと運用コストを削減する高密度1Uサーバー

- 最大16コア/32スレッドのインテル® Xeon® プロセッサーE5-2600製品ファミリーを搭載、24DIMMスロット、4x1GbEまたは2x10GbEのNICを装備し、仮想化プラットフォームとしても最適
- 最大10台の2.5型ハードディスクを1Uサイズに内蔵可能、1Uでも余裕のストレージ容量
- iLOマネジメントエンジン(iLO4)がサーバーのライフサイクル管理を自動化
- 「センサーの海」により、冷やすべき箇所だけ冷やすスマートな冷却を実現、独自の省電力機能を継承しつつさらに高性能となった電源を装備し、無駄な電力消費を徹底して排除

キーテクノロジー3

Red Hat Enterprise Linux

企業向けLinuxとして抜群の評価と採用実績

オススメ構成例のOSとしてHPが採用したRed Hat Enterprise Linuxは、オープンソースをベースとし、企業向けLinuxのグローバルスタンダードといえる存在です。ビッグデータやHPCに対応したスケールアウトによる高速処理、仮想化機能の標準搭載、堅牢なセキュリティなどエンタープライズ規模のシステムニーズに応えます。9000種類を超えるビジネスアプリケーションの稼働実績、基幹系業務の安定稼働も支える安心の企業向けサポート体制も大きな魅力です。

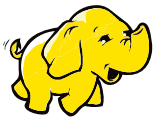
redhat

キーテクノロジー4

Cloudera's Distribution including Apache Hadoop (CDH4)

安定性と使いやすさが認められた標準Hadoop

オープンソースの分散処理システムとして誕生したApache Hadoop。その中心的な開発者であるダグ·カッティングの所属するベンチャー企業Clouderaが提供するCDH4は、Apache Hadoopプラットフォームの標準ディストリビューションとして世界的な地位を獲得しています。Apache Hadoopプロジェクトで認可されたコードのみを採用した100%Apacheライセンスだから、稼働の安定性は抜群。充実した運用管理ツール群や保守サービスも整備されており、エンタープライズクラスの要請にも応えます。



⚠ 安全に関するご注意　ご使用の際は、商品に添付の取扱説明書をよくお読みの上、正しくお使いください。水、湿気、油煙等の多い場所に設置しないでください。火災、故障、感電などの原因となることがあります。

お問い合わせはHPカスタマー・インフォメーションセンターへ
03-5749-8330 月～金 9:00～19:00　土 10:00～17:00(日、祝祭日、年末年始および5/1を除く)
機器のお見積もりについては、代理店、または弊社営業にご相談ください。
HP Hadoop HBaseサービスに関する情報は www.hp.com/jp/Hadoop_HBase
HP & Red Hat 情報ポータルは www.hp.com/jp/redhat
HP ProLiant Gen8に関する情報は www.hp.com/jp/proliant

Intel、インテル、Intel ロゴ、Intel Inside、Intel Inside ロゴ、Intel Atom、Intel Atom Inside、Intel Core、Core Inside、Intel vPro、vPro Inside、Celeron、Celeron Inside、Itanium、Itanium Inside、Pentium、Pentium Inside、Xeon、Xeon Inside、Ultrabook は、アメリカ合衆国および/またはその他の国における Intel Corporation の商標です。
記載されている会社名および商品名は、各社の商標または登録商標です。
記載事項は2012年10月現在のものです。
本カタログに記載されている情報は取材時におけるものであり、閲覧される時点で変更されている可能性があります。あらかじめご了承ください。
© Copyright 2012 Hewlett-Packard Development Company,L.P.

本カタログは、環境に配慮した用紙と植物性大豆油インキを使用しています。 PRINTED WITH SOY INK

日本ヒューレット·パッカード株式会社
〒136-8711 東京都江東区大島2-2-1

JSS12445-01